DOCUMENT RESUME

ED 379 343 TM 022 730

AUTHOR Tasse, Marc J.; And Others
TITLE Construction of a Computerized Adaptive Testing Version of the Quebec Adaptive Behavior Scale.
PUB DATE Jun 94
NOTE 11p.; Paper presented at the Annual Meeting of the American Association on Mental Retardation (118th, Boston, MA, May 31-June 4, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Ability; *Adaptive Testing; *Adjustment (to Environment); Adults; Behavior Rating Scales; *Computer Assisted Testing; Databases; *Estimation (Mathematics); Item Banks; Item Response Theory; Mental Retardation; Simulation; Statistical Analysis; *Test Construction; Test Items
IDENTIFIERS Item Parameters; *Quebec Adaptive Behavior Scale

ABSTRACT

Multilog (Thissen, 1991) was used to estimate parameters of 225 items from the Quebec Adaptive Behavior Scale (QABS). A database containing actual data from 2,439 subjects was used for the parameterization procedures. The two-parameter-logistic model was used in estimating item parameters and in the testing strategy. MicroCAT (Assessment Systems Corporation, 1989) was then used to manage the item banks and Computerized Adaptive Testing (CAT) environment during a simulation run using data from a randomly selected sample of 200 subjects taken from the larger data base. The simulation of the QABS-CAT testing indicates that levels of ability can be estimated for each of the seven skill domains by using only 30% of the items of the conventional version. The numerous advantages of item response theory and CAT as applied to the assessment of adaptive behavior with regard to the changing definition of mental retardation (Luckasson et al., 1992) are discussed. (Author/SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

9118

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MARC J. TASSÉ

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

ED 379 343

# Construction of a Computerized Adaptive Testing Version of the Quebec Adaptive Behavior Scale

Marc J. Tassé

Nisonger Center, The Ohio State University

Paul Maurice, & Réjean Auger

Université du Québec à Montréal

Running head: Construction of the QABS-CAT

TM022730



BEST COPY AVAILABLE

Abstract

Multilog (Thissen, 1991) was used to estimate the parameters of the 225 items from the QABS. A database containing actual data from 2439 subjects was used for the parameterization procedures. The two-parameter-logistic model was used in estimating item parameters and in the testing strategy. MicroCAT (Assessment Systems Corporation, 1989) was then utilized to manage the item banks and CAT environment during a simulation run using data from a randomly selected sample of 200 subjects taken from the larger data base (N=2439). The simulation of the QABS-CAT testing indicates that levels of ability can be estimated for each of the seven skill domains by using only 30% of the items of the conventional version. The numerous advantages of IRT and CAT as applied to the assessment of adaptive behavior with regards to the changing definition of mental retardation (Luckasson *et al.*, 1992) are discussed.

The AAMR's 9th manual on definition, classification, and systems of supports of mental retardation (Luckasson *et al.*, 1992) operationally defines adaptive behavior as consisting of ten specific adaptive skills. The DSM-IV has integrated this aspect of the AAMR definition into its diagnostic criteria for mental retardation (APA, 1994). The operationalization of adaptive behavior into ten specific adaptive skills has created a need for new instruments in the area of adaptive behavior measurement (MacMillan, Gresham, & Siperstein, 1993).

This paper explores the possibilities of exploiting some of the work coming out of the Educational Testing Service with regards to Item Response Theory (IRT) in the construction of a computerized adaptive testing version of the Quebec Adaptive Behavior Scale (QABS; Maurice, Morin, & Tassé, 1993).

IRT is an alternative testing theory to classical testing theory and plays a crucial role in Computerized Adaptive Testing (CAT), also called tailored testing. It is expected that CAT will permit the estimation of adaptive behavior functioning within a predetermined error of measurement with significantly fewer items presented than the conventional QABS.

## Method

### Subjects

#### Calibration

The data from 2,439 subjects contained in an accumulated database were used to estimate the item parameters. It is important to note that these are real data and thus represent optimal parameter estimation data, as opposed to computer generated (Monte Carlo) data.

#### Simulation

A randomly selected sample of 200 subjects from the initial pool of the 2,439 subjects

was used to complete a simulation of the QABS-CAT version.

Procedure

The initial step of this project involved the calibration of the item banks using a two-parameter-logistic model. This model allowed the estimation of the item difficulty ($b_i$) and item discrimination ($a_i$) from the database containing 2,439 subjects. The score vectors of the 2,439 subjects to the 225 items (part one) of the QABS were imported into a data file to be interpreted by Multilog (Thissen, 1991).

A simulation was conducted rather than using live evaluators so as not to confound the human-computer interaction (Mathisen, Evans, Meyers, & Kogan, 1985) factor in the assessment procedure using the QABS-CAT. The objective of this simulation was to estimate the average number of items needed in adaptive strategy employed by the QABS-CAT to estimate ability level for each of the seven skill domains. The data for the 200 subjects were entered manually by the first author and four graduate students.

Testing strategies.

The Bayesian testing strategy was privileged within the adaptive testing strategy. The items are selected on a maximum information basis which roughly translates in the preference for more discriminating items over less discriminating items when the $b_i$ are equal. The test (per domain) is terminated when either of the two stopping conditions are met. The first stopping rule contains two conditions which are: a) the error variance surrounding the $\Theta$ estimate is inferior to 0.1 (which represents a 95% confidence interval) and b) a minimum of six-items have been presented. The second stopping rule, if the first condition is never met, is the expiry of the item bank. If no items are remaining in the bank, the test moves on to the next item bank or stops the

testing session if all domains have been assessed, even if the error variance surrounding the Θ estimate has not met the stopping criterion.

## Results

The difficulty levels for the seven domains of the QABS-CAT for a probability of .50 and the point along ability level (Θ) where information function [I(Θ)] peaks are presented in Table I. The I(Θ), which represents the degree of precision of the Θ estimates for the domain, are also presented in Table I.

Figure 1 graphically represents the distribution and frequency along $b_i$ for individual items in each of the seven domains. A test, be it adaptive or conventional, is only as strong as the item bank from which it draws its items for Θ estimates. The information presented in Figure 1 permits us to quickly glance at areas where certain domains are lacking items of certain difficulty levels.

### Simulation data

The simulation permitted us to verify the proper functioning of the QABS-CAT testing strategy. Of the 200 subjects used in the simulation of the QABS-CAT, the average number of items needed to estimate Θ levels for the seven domains was 67 items (30% of QABS) (Figure 2), with a range between 42 (19% of QABS) and 102 items (45% of QABS). The simulation data indicate that the QABS-CAT was able to complete its ability estimates for the simulation sample with an average of 30% of items from the original QABS item bank, representing a net economy in time and item presentation. The subjects who were presented the highest number of items tended to be subjects with either the highest or lowest Θ estimates.

## Conclusion

The simulation using 200 randomly selected subjects confirmed the practicability of the QABS-CAT. The objective with this simulation was to ascertain the proper running of the programming as well as to determine the number of items required to estimate $\Theta$ levels within the predetermined error variance. The results of the simulation confirmed the net economy in item presentation, hence economy in administering time. On average only 30% of the item bank was required to accurately estimate the ability levels for the seven skill areas.

The fundamental advantages of the CAT is the individualized measurement of adaptive behavior and error estimate surrounding each ability level. For every subject assessed, the error estimate is calculated and the test terminates only if this error variance is below 0.1, which represents a 95% confidence interval.

Adaptive behavior is a complex construct to measure and requires large item banks, which makes CAT an interesting avenue to explore. Being an age indexed construct only augments the need for tailored testing, a possibility offered through CAT. This project was a first attempt to apply IRT and CAT to the measurement of adaptive behavior. In light of Luckasson *et al.* (1992) and DSM-IV (APA, 1994), IRT and CAT could prove to be a valuable tool permitting reliable and efficient ability estimation of the ten specific adaptive skills of adaptive behavior.

## References

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th edition). Washington, DC: Author.

Luckasson, R., Coulter, D.L., Polloway, E.A., Reiss, S., Schalock, R.L., Snell, M.E., Spitalnik, D.M., & Stark, J.A. (1992). Mental retardation: Definition, classification, and systems of supports (9th ed.). Washington, DC: AAMR.

MacMillan, D.L., Gresham, F.M., & Siperstein, G.N. (1993). Conceptual and psychometric concerns about the 1992 AAMR definition of mental retardation. American Journal on Mental Retardation, 98, 325-335.

Mathisen, K.S., Evans, F.J., Meyers, K., & Kogan, L. (1985). Human factors influencing patient-computer interaction. Computers in Human Sciences, 1, 163-170.

Maurice, P., Morin, D., & Tassé, M.J. (1993). Manuel technique de l'Échelle québécoise de comportements adaptatifs. Montréal, QC: Université du Québec à Montréal.

Thissen, D. (1991). Multilog: User's guide (version 6.0). Chicago, IL: Scientific Software Incorporated.

**Table I**

TCC and information function

| Domain | Difficulty $b_i$ | Maximum information $I(\Theta)$ |
|---|---|---|
| Personal Autonomy | -0.8 | -0.9 |
| Domestic Skills | 0.8 | 0.6 |
| Health and Sensory-motor | -0.8 | -1.2 |
| Communication | -0.5 | -0.7 |
| Academic Skills | 1.3 | 1.2 |
| Social Skills | 0.4 | -0.1 |
| Vocational Skills | 0.8 | 0.5 |

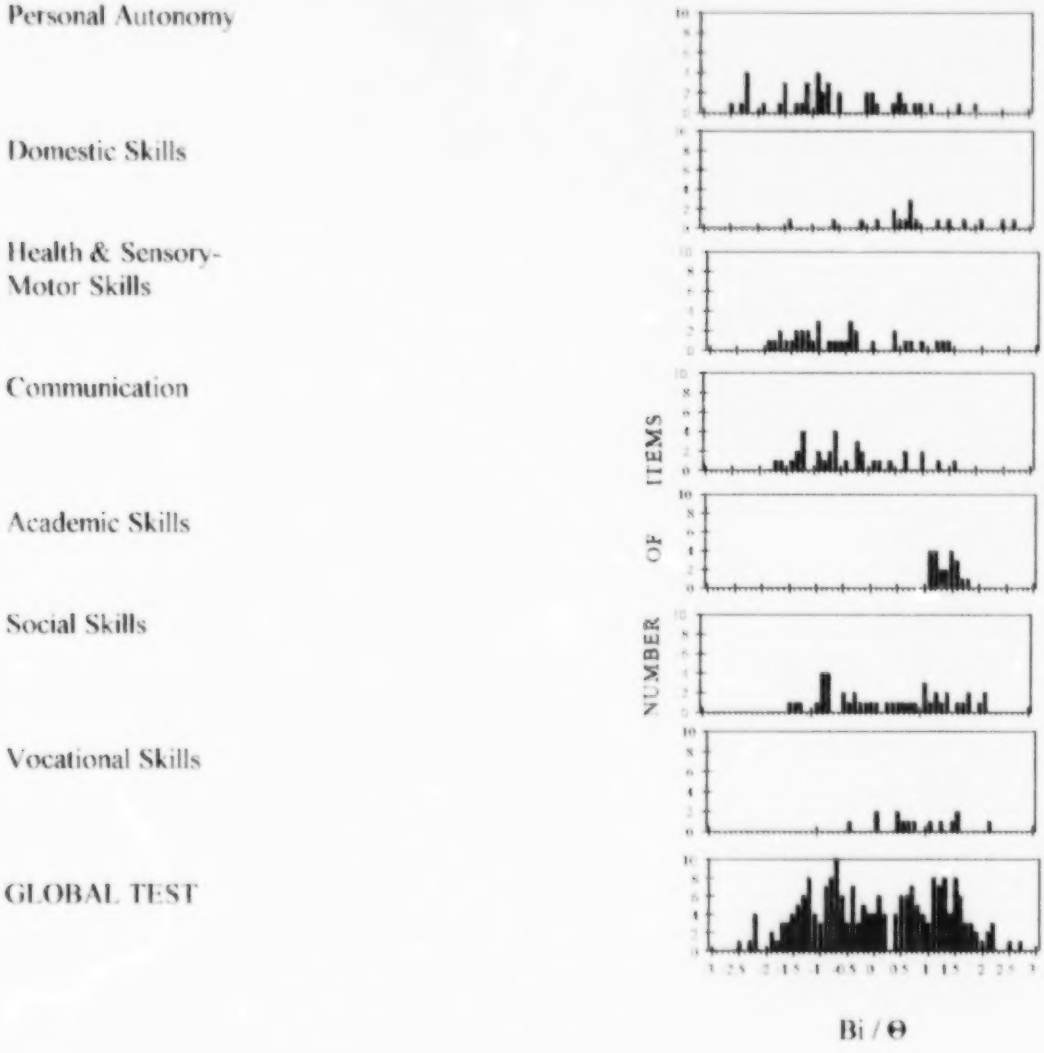


Figure 1 Number of items along the $B_i / \Theta$ continuum for all seven skills domains and global test

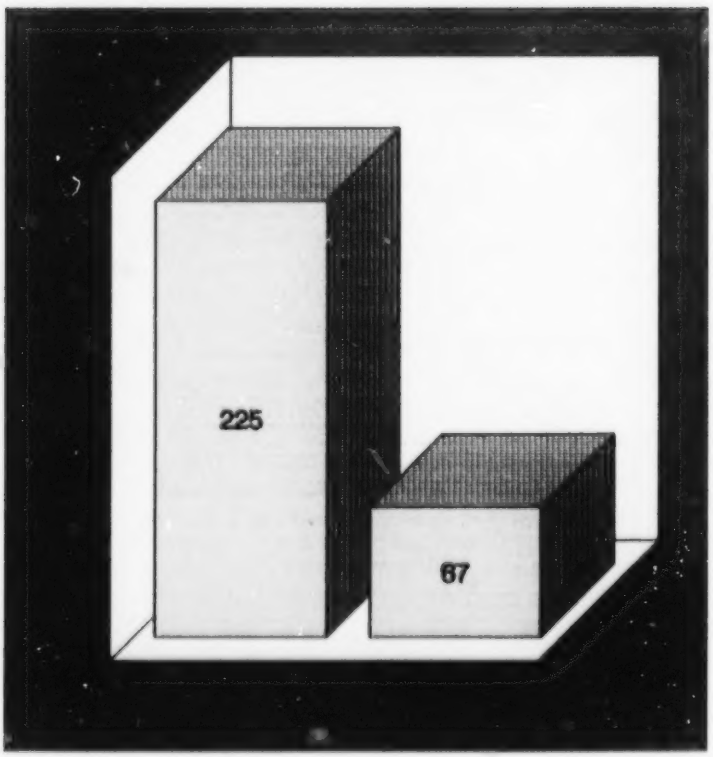


Figure 2 Average Total Number of Items Presented QABS versus QABS-CAT.